`U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
OFFICE OF SCIENCE AND TECHNOLOGY
METEOROLOGICAL DEVELOPMENT LABORATORY

MDL OFFICE NOTE 12-1

BIAS CORRECTION OF MOS TEMPERATURE AND DEWPOINT FORECASTS

Bob Glahn

December 2012

BIAS CORRECTION OF MOS TEMPERATURE AND DEWPOINT FORECASTS

Bob Glahn

1. INTRODUCTION

There has been considerable discussion in the literature over the past several years concerning short-sample bias correction of model and MOS forecasts. Also, bias correction is being done at many National Weather Service (NWS) Weather Forecast Offices (WFO). Much of the discussion is related to ensembles, but generally the methods employed are applied to each member individually. For a series of forecasts of a weather element, such as 2-m temperature 48 hours in the future, and a measure of "truth," call it an observation, at the valid times of the forecasts, a series of forecast errors can be created by subtracting each observation from the forecast. According to the usual definition of bias used in meteorology (Wilks 2011, p. 304; Jolliffe and Stephenson 2003, pp. 99-100; Murphy and Daan, p. 385 in Murphy and Katz 1985), if the average of a consecutive set of errors is sufficiently different from zero, the forecasts would be considered to be biased. Most of the authors discussing bias correction do not specifically state over what period of time they would consider the forecasts to be biased.[1] The term "bias correction" does not, in general, refer to correcting the bias, but rather in correcting individual forecasts based on the recent past bias of those uncorrected forecasts. It is noted that such a correction method cannot respond quickly to errors of forecasts X days into the future, because the error cannot be calculated with the current observation until X days have passed.

Weather elements for which bias correction is discussed are mainly "continuous" in time and space, such as temperature, as opposed to those of a more discontinuous nature, such as ceiling height. The weather is driven by the sun, and the seasonal cycle is well known. Largely because of the irregular locations, shapes, and elevations of land masses, the yearly changes are far from regular, and to explain the average variability, more harmonics are necessary and can be defined for the variables and locations desired. There is no question, these harmonics in the weather variables, and likely in their forecasts and errors, exist. Even so, they are not necessarily of the same magnitude, or even there at all, year to year, and whether or not they can be predicted to a better degree than can be done with NWP models and postprocessed by traditional MOS is still questioned by some. The studies reported in the literature are predominantly applications to raw NWP output, and not MOS. It is well known the NWP output is inaccurate in various ways, and systematic biases can be corrected; this is one objective of MOS, and it is quite successful when an adequate sample of data is available. The methods applied by the Meteorological Development Laboratory (MDL) and implemented operationally at the National Centers for Environmental Prediction (NCEP) generally deal with 6-month seasons, and for the developmental sample the forecasts have zero bias over the seasons of the sample. Certainly, shorter period biases can and undoubtedly do exist. These biases are likely mostly related to either weather regimes (e.g., blocking highs or persistent storm tracks) that are not regular or of fixed length, or to much longer-scale patterns that are set up (e.g., El Nino and North Atlantic Oscillation) that can be observed but are also not of fixed periods and cannot be forecast well.

[1] Some authors have called a single forecast error a "bias" (Cui et al. 2012, p. ; Cheng and Steenburgh, 2007, p. 1317; Roeger and Stull, 2003, p. 1158) but that terminology can be confusing and does not follow the definition of bias.

A source of MOS error is the inability to "keep up" with the operational model changes. Changes in a model, without redeveloping MOS on an adequate sample, can create larger errors in MOS forecasts.[2] Short-sample bias correction is seen as a way of correcting such biases, and now that NWP models are much better than a few years ago, possibly bias-corrected raw NWP 2-m temperature is of comparable accuracy to MOS temperature. However, Cheng and Steenburgh (2007) found that in 2003 MOS forecasts were better than bias-corrected model forecasts of temperature except in periods of "quiescent large-scale patterns."

Quite a number of bias-correction methods have been devised and tested (e.g., Yussouf and Stensrud 2007; Woodcock and Engel 2005; Fan[3]), and most show some improvement over raw NWP or MOS forecasts. Testing within MDL of three methods has shown little difference among them for accuracy and low bias of forecasts. A method used at NCEP is called "decaying average." It is perhaps the simplest of correction methods and very easy to implement. Simply, a forecast is corrected by applying a "delta" that is a weighted average of the error of the forecast most recently verified and the weighted average of previous deltas. The average "decays" with time so that errors in the distant past are discounted more than the recent ones. Response curves have been shown by Cui et al. (2012). All that is not known is the weight α to apply to the current error and $(1-\alpha)$ to apply to the past average error. This paper addresses the selection of α based on the errors and biases of resulting bias-corrected forecasts, and the arguments for and against implementation. An appendix explains the software that was used to do the study, and that could be used to implement the method within the MDL MOS system.

2. DECAYING AVERAGE ALGORITHM

Cui et al. (2012, abstract) call the decaying average implemented at NCEP in 2006 a "Kalman filter type algorithm," but do not state the assumptions made in deriving it. The use of the Kalman filter as originally proposed (Kalman 1960) requires estimates of several constants and parameters. Kalman's original work (1960) has been described in several publications in the meteorological literature (e.g., Roeger et al. 2003; Homleid 1995; Galanis and Anadranistakis 2002; Cheng and Steenbaugh 2007) and won't be repeated here. NCEP's algorithm is very attractive because is very easy and cost effective to implement.

To implement the algorithm, one has only to carry forward a delta and apply it to the current forecast. Then to prepare for the next forecast cycle, the delta d would be updated by:

$$d\,(t+1) = (1-\alpha)\; d\,(t) + \alpha\,(F - O)(t)$$

where $d(t+1)$ is the delta to apply at time $t+1$, $d(t)$ is the delta applied at time t, F is the forecast "verified" by the observation O at time t, and α is the weight to apply to the most recently calculated forecast error $F - O$ at time t. There would optimally be a specific delta for each station and forecast projection. When $F - O$ is missing, assume zero (see the appendix for more discussion).

---

[2] A *major* improvement to the model would likely decrease the MOS errors, but such major improvements rarely happen; improvement comes in small steps.

[3] Unpublished briefings in 2012.

3. PERFORMANCE OF THE ALGORITHM APPLIED TO MOS FORECASTS

Before implementing bias correction into the MOS system, several questions have to be addressed, including not only whether the accuracy and bias of the forecasts are improved, but also what effect this might have on the customers and partners who use the "forecasts" directly or manipulate them by either automated or manual means to achieve a "final" forecast. In the NWS, this further processed guidance would be the "official" forecast. These questions are addressed in this paper.

A. Data Used

I had available operational GFS-based MOS temperature and dew point forecasts made at 0000 UTC over the period January 1, 2011, through May 31, 2012,[4] for projections every 12 hours out to 264 hours (11 days). This provided a sufficient sample on which to investigate the bias correction method. A change was made to the Global Forecast System (GFS) at NCEP during the summer of 2012 that had some major negative effects on MOS in some portions of the United States, but primarily to wind, not temperature.

Most testing was done separately on the traditional MDL MOS warm (March-September) and cool (October-March) seasons, although for dewpoint, all data were processed together. When comparing forecasts of different projections, one can either match verification times, the forecasts having been made previously at different times, or one can verify the forecasts based on when they were made, and the period over which observations were used will not exactly match for different projections. The latter is the way it is usually done and the way I carried out the verification. Evaluated over 6-month or longer seasons, the offset in verification times is not a significant factor in conclusions for projections up to 11 days. The comparison of MOS forecasts and bias corrected (BC) forecasts for a particular projection were, of course, matched samples. The verification was done with forecasts rounded to whole degrees Fahrenheit, the same precision as the observations. The deltas, however, were carried to three decimal places.

B. Performance of Different Alphas

<u>Bias</u>

NCEP and MDL have investigated different alphas up to at least 0.1; NCEP uses 0.02 for all projections for raw model data[5]; testing at MDL had indicated a higher value might be better for MOS. I tested four different values: 0.025, 0.05, 0.075, and 0.1. I used 1319 stations in the conterminous states (CONUS), the same set used in routine MDL verifications. Of course, a few observations, and even forecasts, may be missing in the sample. The adjustment algorithm can deal with occasional missing data; if there is no forecast or matching observation for a particular projection, no change in the delta is necessary, and on the next cycle for which there is

---

[4] Being operational forecasts, the dew point forecasts had been checked with temperature forecasts, and if a forecast was greater than the temperature, it was set to the temperature. If bias correction were implemented, the temperature/dewpoint check would come *after* the bias correction. Using the checked values for this study instead of unchecked ones is not seen as a problem, because the number and magnitude of the changes are rather small.

[5] Cui 2012 personal communication.

a forecast, the delta last calculated is used. However, a station may miss reporting for an extended period or stop altogether. MOS forecasts will likely still be made because the model data are available. To perpetuate a particular bias indefinitely would not be prudent, so a better alternative is to use the difference as zero and let the average decay gradually to zero. That is, given no recent history of the station's bias, a correction is not made. In either case, no elaborate backup software is necessary to accommodate one or more missing pairs of data. For the testing reported here, the decay toward zero was not used.

Figure 1 shows the bias for the original MOS temperature and the BC forecasts with all four alphas tested for all stations for all projections for the 2011-2012 cool season. It can be seen that there was a significant cold bias that varied each 12 hours and generally was more pronounced with increasing projection. Forecasts verifying at some times of the day have larger errors than forecasts verifying at other times of the day; that is what causes the saw-tooth effect. It can also be seen that the BC forecasts are better in terms bias, being positive but near zero at 24 hours, but still have a cold bias approaching 0.05 degrees F at later projections. It is also interesting that for the early projections, the projections with the largest negative bias for MOS have the largest positive bias for BC MOS. This mirror performance lasts until about 228 hours, when it reverses and the two curves come into phase. The bias was improved substantially for all alphas. The largest improvement was for alpha = 0.1 and the smallest for 0.025 for all projections, but the differences are small, especially compared to the MOS bias.

The bias varies considerably over the CONUS. Figure 2 is the same as Fig. 1 but for the NWS Central Region, the region for which the bias is greatest, and similarly, Fig. 3 shows biases for the Western Region, the region for which the bias is least (note the different ordinate scales in these figures). Fig. 2 shows a remarkably negative bias of 3 deg. F at 264 hours; the BC forecasts are for every projection better, but still drift systemically down to -0.06 degrees F at 264 hours for alpha = 0.05. On the other hand, the biases for the Western region range generally from -.05 to +.08 degrees, and the BC forecasts are better for most projections, especially for the larger alphas.

Figure 4 is the same as Fig. 1 except for the 2011 warm season, April through September. The pattern of MOS error by projection is dissimilar to the cool season, and the improvement is questionable except for projections $\leq$ 72 h. For projections of 84 hours and beyond, the MOS bias was very small, and the correction was not, in general, helpful. As with the cool season, alpha = 0.1 was the best and 0.025 the least helpful.

Rather than separate the dewpoint data into seasons, the whole sample January 2011 through May 2012 was used. For each test done, the process was "cold-started" with a delta = 0 on January 1. Because of this, the verification period for dewpoint started on January 15, allowing a stabilizing period. Figure 5 shows the results. Again, MOS bias is improved, and all alphas produce a very small positive bias. Here, the small values of alpha were best, but the differences are minuscule. Probably one reason the biases of the corrected forecasts do not vary much and are so close to zero is because of the long averaging period over all 16.5months.

Mean Absolute Error

Figures 6 and 7 show mean absolute errors (MAE) for all stations for MOS and the four values of alpha tested for temperature for the cool and warm seasons, respectively. Only forecasts verifying at 0000 UTC are shown. Forecasts verifying at 1200 UTC show the same pattern, but the errors are considerably larger and if both are shown on the same graph, the saw-tooth pattern presents a less clear picture. One might hope with the improvement in bias, that the MAEs would also improve (decrease). Indeed they do, but not with all alphas. For the lower values of 0.025 and 0.05, there is consistent improvement, but with the two higher values of 0.075 and 0.1, there are larger MAEs in the cool season. Because the improvement with the smaller alphas is consistent for all projections, especially for the warm season, and `paired t-tests for each projection show very high significance,[6] some slight improvement can be expected in the future.

Figure 8 is similar to Figs. 6 and 7, except for dewpoint. The conclusion for dewpoint is the same as for temperature; the improvement with bias correction is substantial, on the order of 0.3 degrees F at lower projections; improvement is small but consistent at higher projections. The improvement does not vary much with alpha, but the lower values are better at higher projections.

Small and Large errors

Figures 9 and 10 show the percentage of small errors, those < 5 degrees F, for temperature for the cool and warm seasons respectively. These small errors are more frequent with all alphas, but the smaller alphas give slightly better results. The improvement is roughly equal to a 24-h improvement for the warm season. The same general conclusion is reached from Fig. 11, which is for dewpoint; the relative frequencies of small errors are higher for BC forecasts than for MOS, and lower alphas are slightly better.

Figures 12 and 13 are similar to Figs. 9 and 10, except they are for large errors, those > 15 degrees. There are very few such errors for the short projection times, and reach 1 percent at 11 days for the cool season. For that season, alpha = .025 is the only one that actually has lower errors than MOS, but 0.05 does not have more. For the warm season, alpha = 0.1 is the only one that does not improve on MOS for the higher projections, with the 0.025 and 0.05 being of about equal value according to this score.

Figure 14 shows the percentage of large errors for dewpoint. Improvement over MOS holds for all projections, but only for alpha = 0.025 and 0.05 at longer projections.

Consistency of Forecasts over Projections

Long-projection forecasts have more error than short-protection forecasts. As the forecasts for a particular verifying time are improved with time, they should be as consistent as possible, and not "bounce around" from forecast to forecast. The Convergence Score (Ruth et al. 2009) measures the tendency of the longer range forecast to march "consistently" toward the final short range forecast, a higher score being better with a possible maximum of 1.0. Figure 15 and 16

[6] The paired t-tests were corrected for 1st order autoregression (Wilks 2011, p. 147). The errors of the MOS forecasts and the bias corrected forecasts are highly autocorrelated; the MAE pairs are also autocorrelated, but less so. Contributing to the high autocorrelation is the fact the forecasts were rounded to whole degrees Fahrenheit. Even allowing a reduction in degrees of freedom by a factor of 10 for spatial correlation, the paired t-tests still show high significance.

show this score for the four NWS regions and overall for temperature for the cool and warm seasons, respectively. For the cool season, the higher alphas give worse results than MOS for all regions except the Western; the two lower alphas have essentially the same or better scores than MOS. For the warm season, the highest alpha gives worse results than MOS except in the Western Region. The lower alphas are generally the best, the results for 0.025 and 0.05 being essentially indistinguishable.

Figure 17 shows the convergence scores for dewpoint for the warm and cool seasons combined. Only alpha = 0.025 is able to be about as good as MOS for all regions, but alpha = 0.05 is about the same overall. The two higher alphas are consistently worse than MOS.

Bias by Averaging Time

The biases over the sample have been shown previously. However, biases can be present over shorter periods, and may be negative for one period and positive for another period, and cancel out over the sample. In order to see what effect the bias correction has on biases of a shorter time period, running means of forecasts and of observations were computed over a 20-day period–a period long enough someone might consider consistently high (or low) forecasts to be “biased.” Then, the MAEs of these running averages were computed. This measures the biases, both positive and negative, over 20-day periods. Figure 18 shows the results for the warm season temperature. All alphas show improvement over MOS, with the larger values giving the smaller biases, consistent with the biases over the whole samples. The results, not shown, are similar for the cool season, as they are for dewpoint for the cool and warm seasons combined as shown in Fig. 19.

Figure 20 shows the 72-h forecast MAEs with running means of various lengths up to 30 days. The conclusion is the same, regardless of averaging time; short-term local biases are less for BC forecasts than for MOS for all alphas. This means that if a user is interested in mean forecasts, as opposed to daily non-averaged forecasts, bias corrected forecasts are better than uncorrected, and larger alphas, at least up to 0.1, the maximum value tested, are better. The value at 1 day would be exactly the same as shown in Fig. 6 if the sample periods were the same; for averaged forecasts, the ends of the period are slightly different. It is emphasized that the values in Fig. 20 account for the bias being either positive or negative over the x-day period, and are not the same as the overall biases shown in other figures.

Performance on Individual Days

Figure 21 shows the MOS 48-h forecasts for Appleton, Minnesota, and the deltas for the four alphas for the cool season forecasts from February 1, 2011, through November 28, with the warm season forecasts April through September omitted. The alphas operative on March 31 were used on October 1. During March, the forecasts were generally too high. As this period started in late February, the deltas became negative, and for alpha = 0.1 reached a low of -5.0 degrees. For the first few days in October, the MOS forecasts were decidedly too low. The deltas rose rather quickly, and for alpha = 0.1 became as high as +2.5 degrees. But by the time this peak was reached, the period of consistently low forecasts was over, and then a march back toward zero delta began.

The delta for alpha = .025 was much more conservative, never reaching a negative value of more than about 2 degrees. It, too, rose starting October 1, but never peaked above zero. The other alphas performed at an intermediate manner as expected.

4. DISCUSSION OF RESULTS

It is clear from the data presented that the more volatile nature of larger alphas, while improving bias, hurt the accuracy of the MOS forecasts. Also, the deltas (corrections) to apply to the MOS forecasts can vary quite dramatically over a few days with large alphas. It is doubtful that users would welcome this volatility imposed by large alphas. Alpha = 0.025 is quite conservative and is competitive with alpha = 0.05, although 0.05 is slightly better in providing accuracy as judged by some scores for some projections. An alpha in that range, 0.025 to 0.05, seems to be best overall.

5. IMPLEMENTATION DETAILS

When first initiated, the correction process will not have a "decayed average." This is overcome by assuming zero (a "cold start"). Within a few iterations, the process will have stabilized and the average will be close to what it would have been with any other starting value. However, there are a few iterations at first for which the values may not be appropriate. Because MDL uses seasonal equations, there is a question of what average to use on October 1, the start of the cool season, and on April 1, the start of the warm season. One could use the value from March 31 on October 1, but it might not be appropriate, and Fig. 21 is an example of this; the errors switched from positive to negative going from March 31 to October 1.

One option would be to run the correction procedure year-round, and not switch at the seasonal boundaries. However, an average created on September 30 with warm season equations might not be applicable on October 1 for cool season equations.

Another option would be to have an overlap period in the fall and spring during which both sets of equations were run, and the average could be built up during September with cool season equations and be ready to go on October 1.

Because the average temperature is changing most rapidly during the fall and spring months, the change from cool to warm season equations in the spring and the change from warm to cool season equations in the fall can cause a change in forecasts not mirrored by observations. This hiatus should not be made worse with bias correction.

An option that would smooth-out the fall/spring problem due to process changes would be to run both warm and cool season equations for a two month period in the spring and fall centered on April 1 and October 1. The final forecast could be a weighted average of the forecasts from the warm and cool season equations. For instance, the MOS forecast for September 1 would be 1/61 times the result from the cool season equations and 60/61 times the result from the warm season equations. With this result, the bias correction could be run throughout the year with no breaks. This would completely smooth out the seasonal change.

Errors in automated processes occasionally occur, especially when they are data driven. An extremely erroneous forecast or observation could cause a very large change in a delta. To address this possibility, a cap on the error can be imposed. The testing reported here was done with such a cap. The magnitude of the error allowed for a 24-h forecast was 20 degrees; for an 11-day forecast it was 40 degrees, and it varied linearly in between by projection. For instance, if an error of 90 degrees were calculated for an 11-day forecast, likely due to an incorrect observation, it would be capped at 40 degrees–still a sizeable error, but the effect would not be disastrous.

The same software used in testing the bias correction method can be used in implementation with little change. I used temporary IDs for the deltas and the bias corrected forecasts; these may have to be changed. Scripting in an operational setting would, of course, have to be written and implemented. The appendix contains a brief description of the software and its use in the MOS system.

6. CONCLUSIONS

The decaying average algorithm has been tested with alphas ranging from 0.025 to 0.1. It has been found that all alphas improve the bias of MOS forecasts, but only the alphas of 0.025 and 0.05 consistently improve the MAEs. These results hold for both warm and cool season temperature and for dewpoint over a period consisting of both warm and cool seasons. The lower alphas also provide more forecasts of less than 5 degrees F error and less forecasts with >15 degrees error. In addition only the lower alphas improve the consistency of the bias corrected forecasts from the longer range to the shorter range, as shown by the Convergence Score. Plots of the actual performance in terms of the corrections and their volatility indicate that the larger alphas may not be acceptable to the users of MOS forecasts. All of these results lead to the strong indication that bias correction would improve the MOS temperature and dewpoint forecasts when implemented with an alpha in the range 0.025 to 0.05, the exact value not being very important within that range.

ACKNOWLEDGMENTS

I am indebted to Tamarah Curtis for furnishing me data and scripts, and for tutoring me in generating the graphs; to Valery Dagostaro for assistance in retrieving verification data from MDL's verification program; to Stephen Olson and Brian Zeigler for arranging for and setting up a script to access MDL's IBM computers from my PC; to Yun Fan, Eric Engle, and David Rudack for their previous work and for intensifying my interest in this subject, and to Kathryn Gilbert for careful proofreading. This paper is the responsibility of the author and does not necessarily represent the views of the NWS or any other governmental agency.

REFERENCES

Cheng, W. Y. Y., and W. J. Steenburgh, 2007: Strengths and weaknesses of MOS, running-mean bias removal, and Kalman filter techniques for improving model forecasts over the western United States. *Wea. Forecasting*, **22**, 1304-1318.

Cui, B., Z. Toth, Y.Zhu, and D. Hou, 2012: Bias correction for global ensemble forecast. *Wea. Forecasting*, **27**, 396-410.

Galanis, G., and M. Anadranistakis, 2002: A one-dimensional Kalman filter for the correction of near surface temperature forecasts. *Meteorol. Appl.*, **9**, 437-441.

Glahn, H. R., and J. P. Dallavalle, 2000: MOS-2000. MDL Office Note 00-1. National Weather Service, NAA, U.S. Department of Commerce, Chapter 4, 14 pp.

_____, K. Gilbert, R. Cosgrove, D. P. Ruth, and K. Sheets, 2009: The giddying of MOS. *Wea. Forecasting*, **24**, 520-529.

Homleid, M., 1995: Diurnal corrections of short-term surface temperature forecasts using the Kalman Filter. *Wea. Forecasting*, **10**, 689-707.

Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification, A Practitioner's Guide in Atmospheric Science*. Wiley, 240 pp.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME. J. Basic Eng*. **82**, 35-45.

Murphy, A. H., and R. W. Katz, 1985*: Probability, Statistics, and Decision Making in the Atmospheric Sciences*, West view Press, 545 pp.

Roeger, C., R. Stull, D. McClung, J. Hacker, X. Deng, and H. Modzelewski, 2003: Verification of mesoscale numerical weather forecasts in mountainous terrain for application to avalanche prediction. *Wea. Forecasting*, **18**, 1140-1160.

Ruth, D. P., B. Glahn, V. Dagostaro, and K. Gilbert, 2009: The performance of MOS in the digital age. *Wea. Forecasting*, 24, 504-519.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences, 3rd Edition*. Elsevier, Inc., 676 pp.

Woodcock, F. And C. Engel, 2005: Operational consensus forecasts. *Wea. Forecasting*, **20**, 101-111.

Yussouf, N., and D. J. Stensrud, 2007: Bias-corrected short-range ensemble forecasts of near-surface variables during the 2005/06 cool season. *Wea. Forecasting*, **22**, 1274-1286. .

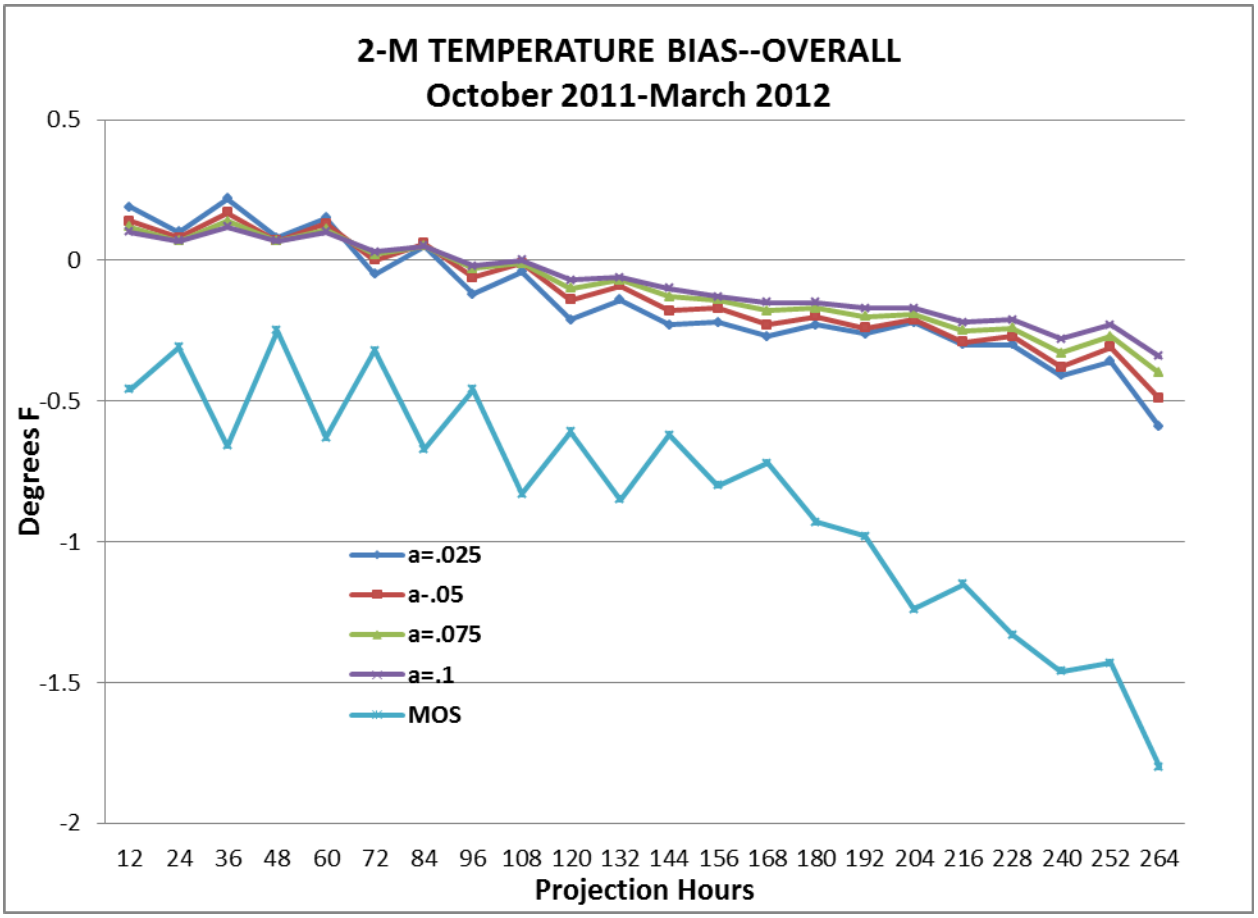


Figure 1. The bias over the 2011-2012 cool season for MOS temperature forecasts and the MOS bias corrected forecasts for the four alphas tested.

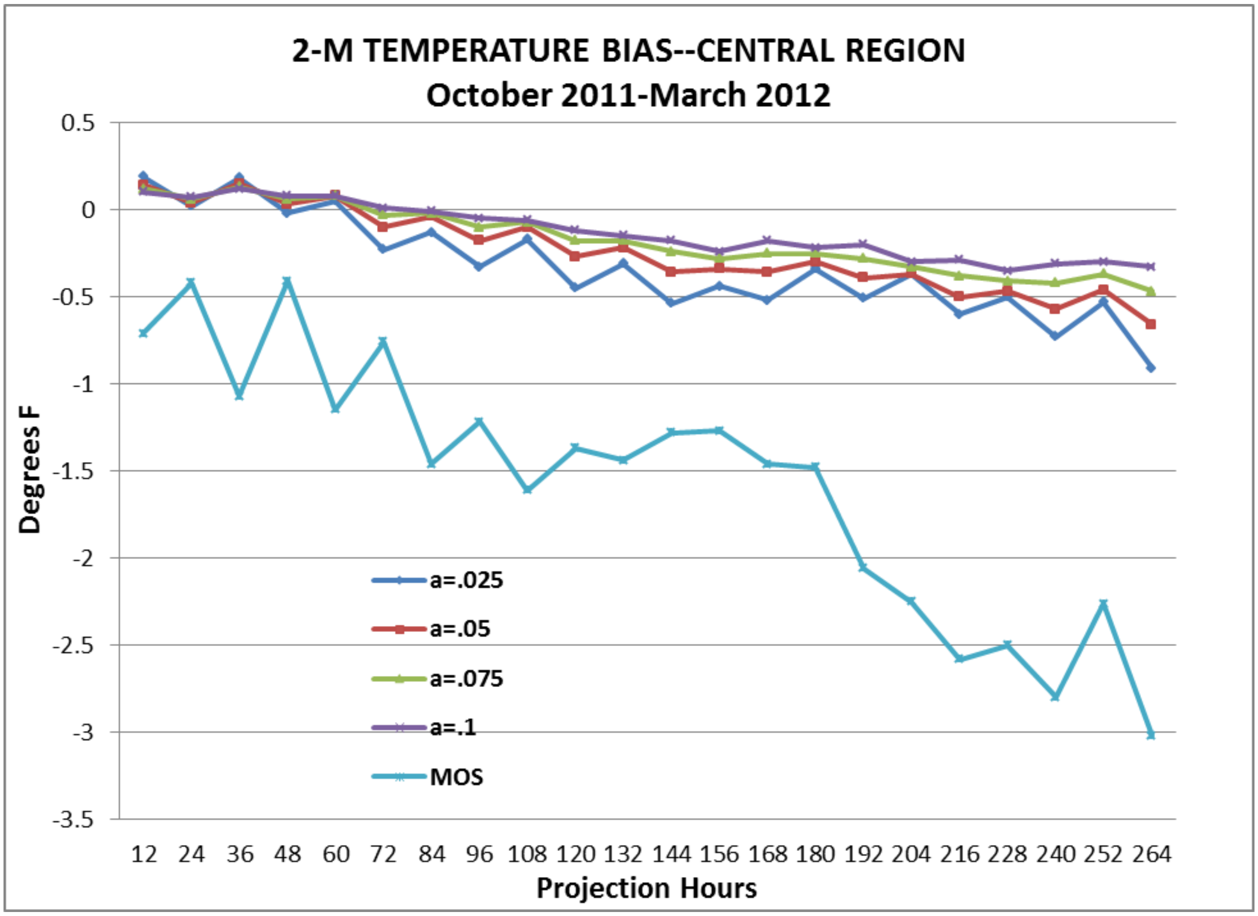


Figure 2. The same as Fig. 1, except for the NWS Central Region.

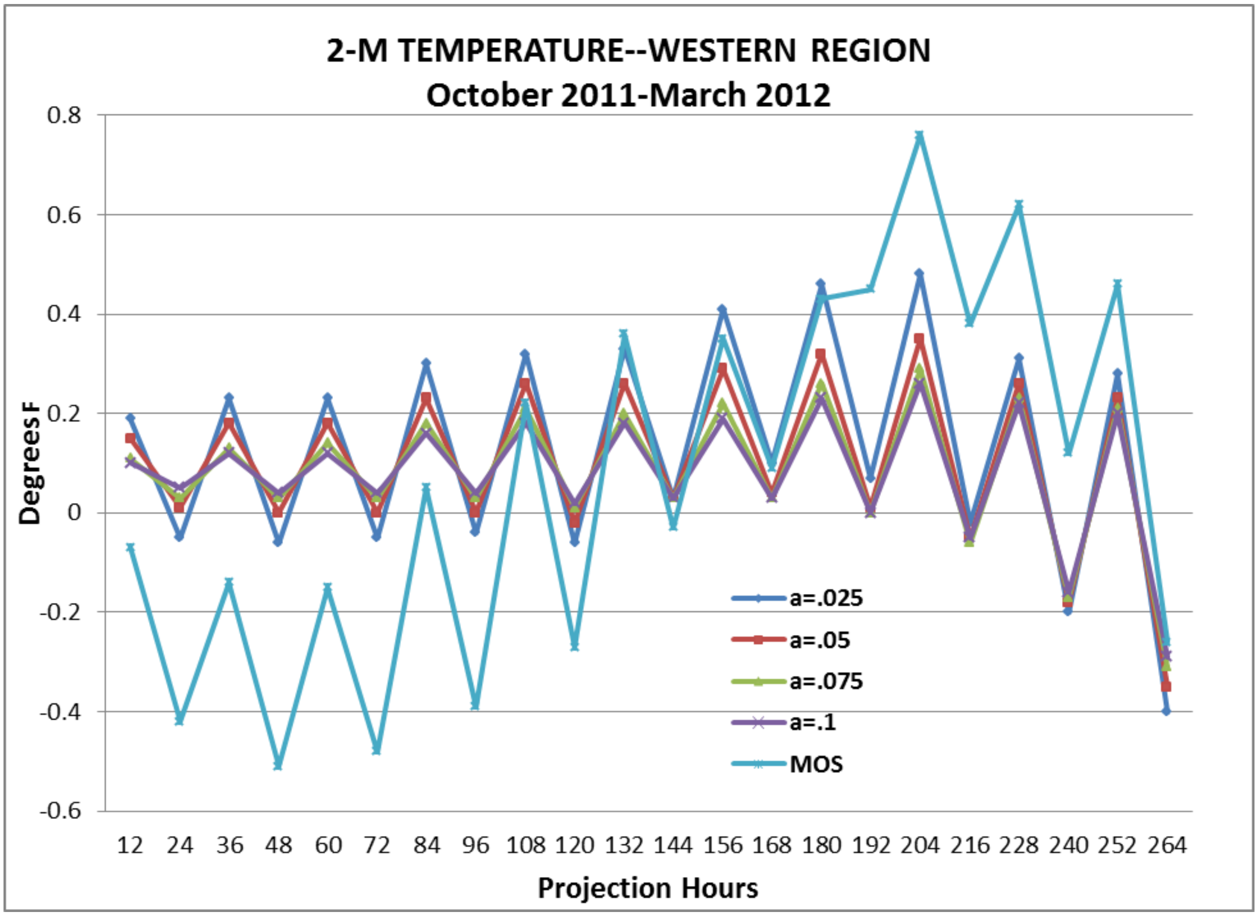


Figure 3. The same as Fig. 1, except for the NWS Western Region.

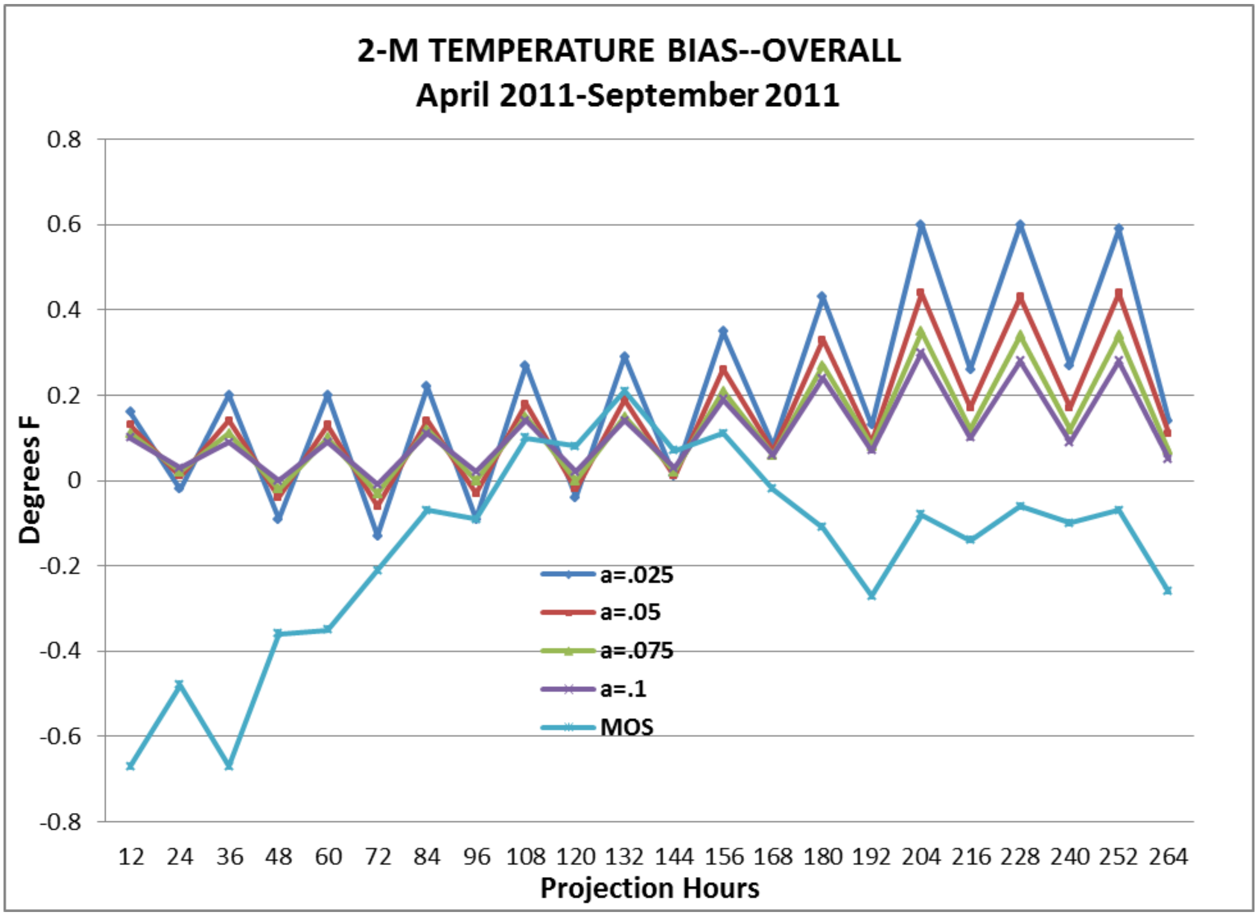


Figure 4. The same as Fig. 1, except for the 2011 warm season.

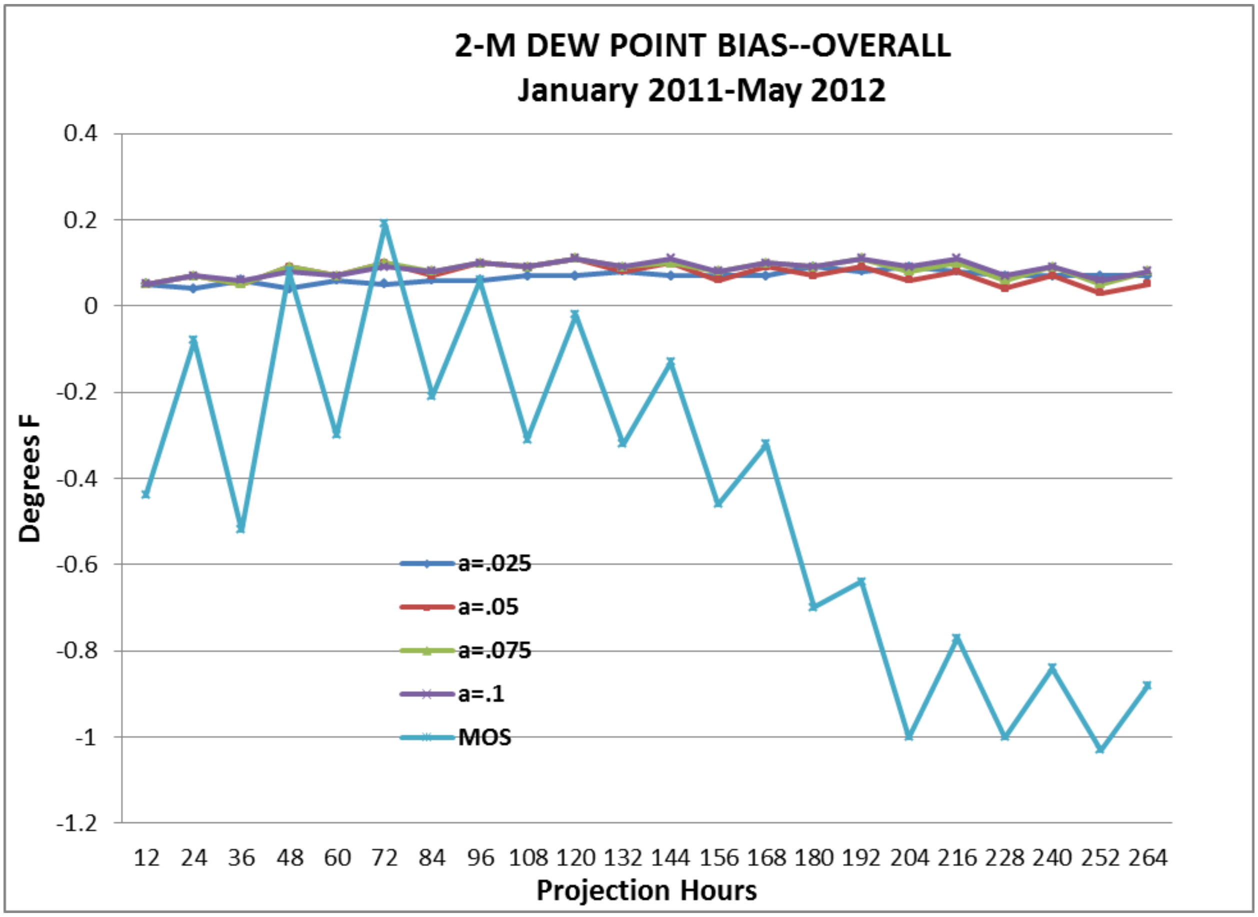


Figure 5. The same as Fig. 1, except for warm and cool seasons combined for dew point.

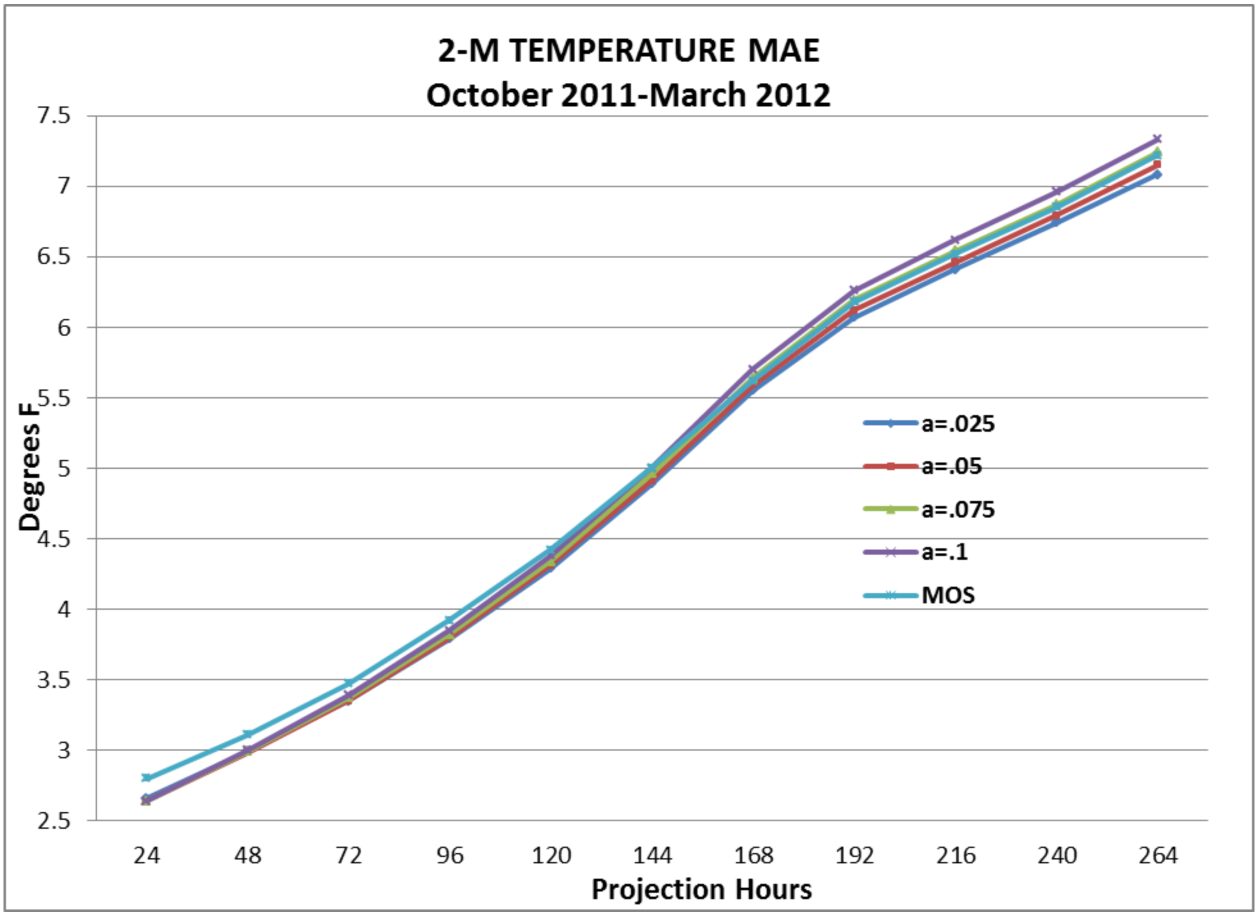


Figure 6. The MAEs over the 2011-2012 cool season for MOS temperature forecasts and the MOS bias corrected forecasts for the four alphas tested. Only forecasts verifying at 0000 UTC are shown.

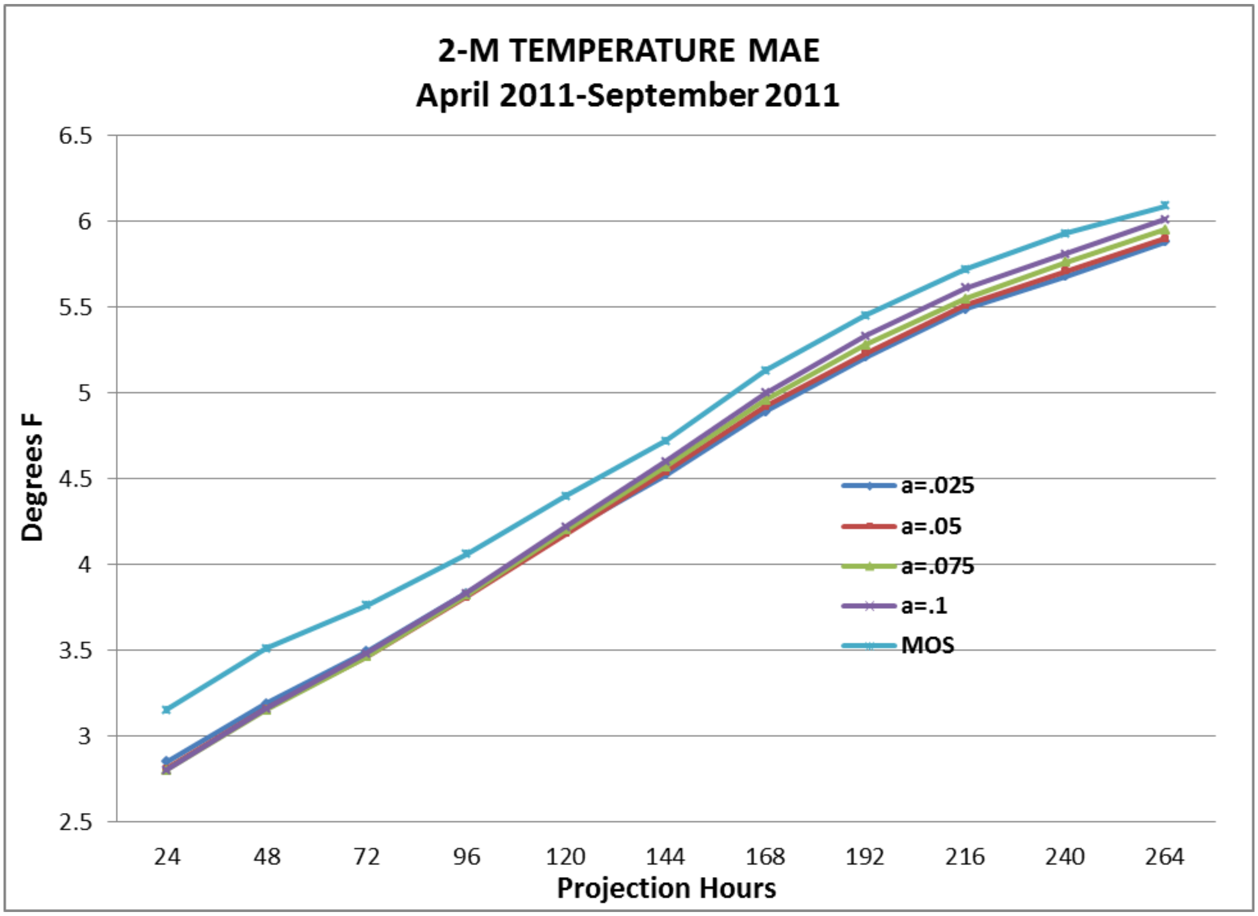


Figure 7. The same as Fig. 6, except for the 2011 warm season.

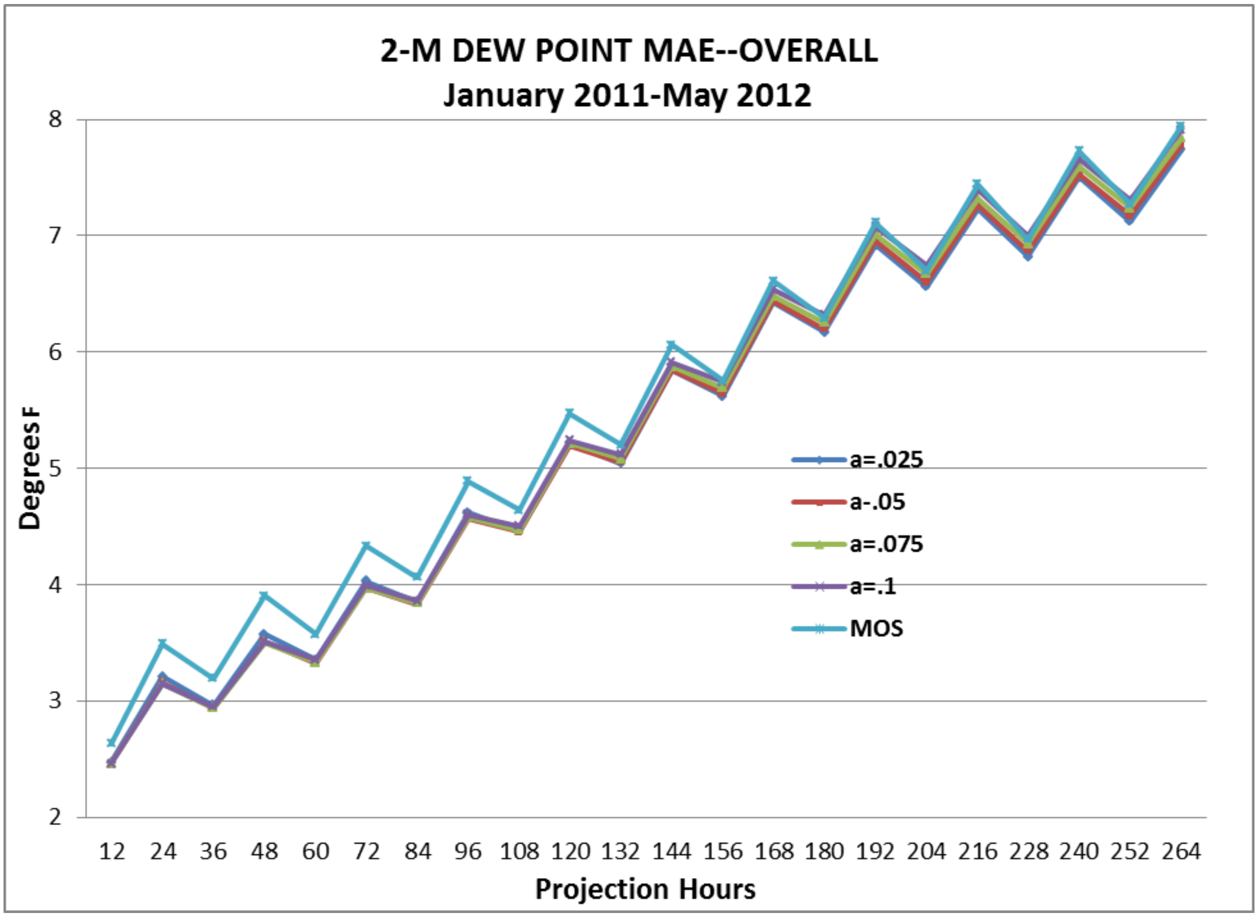


Figure 8. The MAEs over the warm and cool seasons for MOS dew point forecasts and the MOS bias corrected forecasts for the four alphas tested.

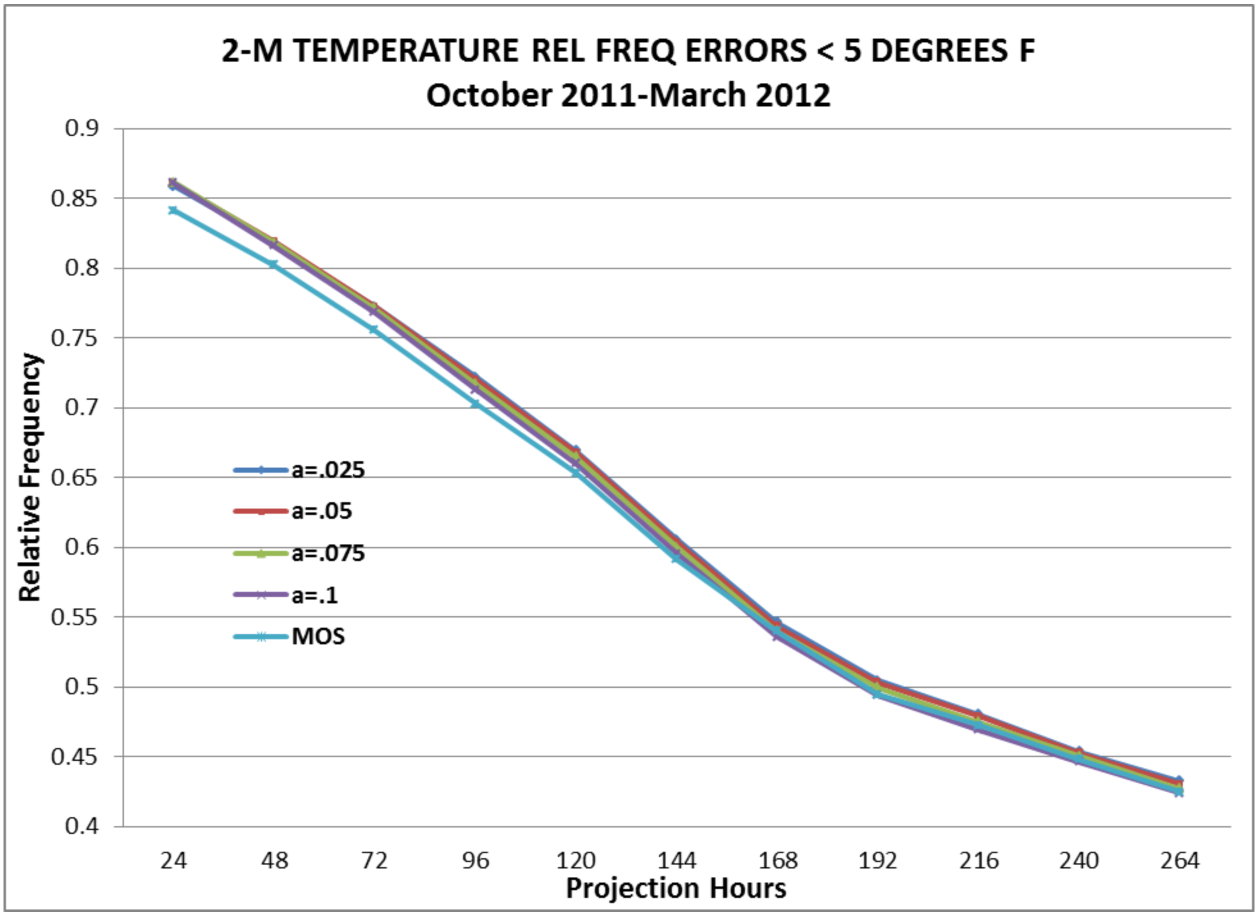


Figure 9. The percentage of temperature errors < 5 degrees F for the cool season.

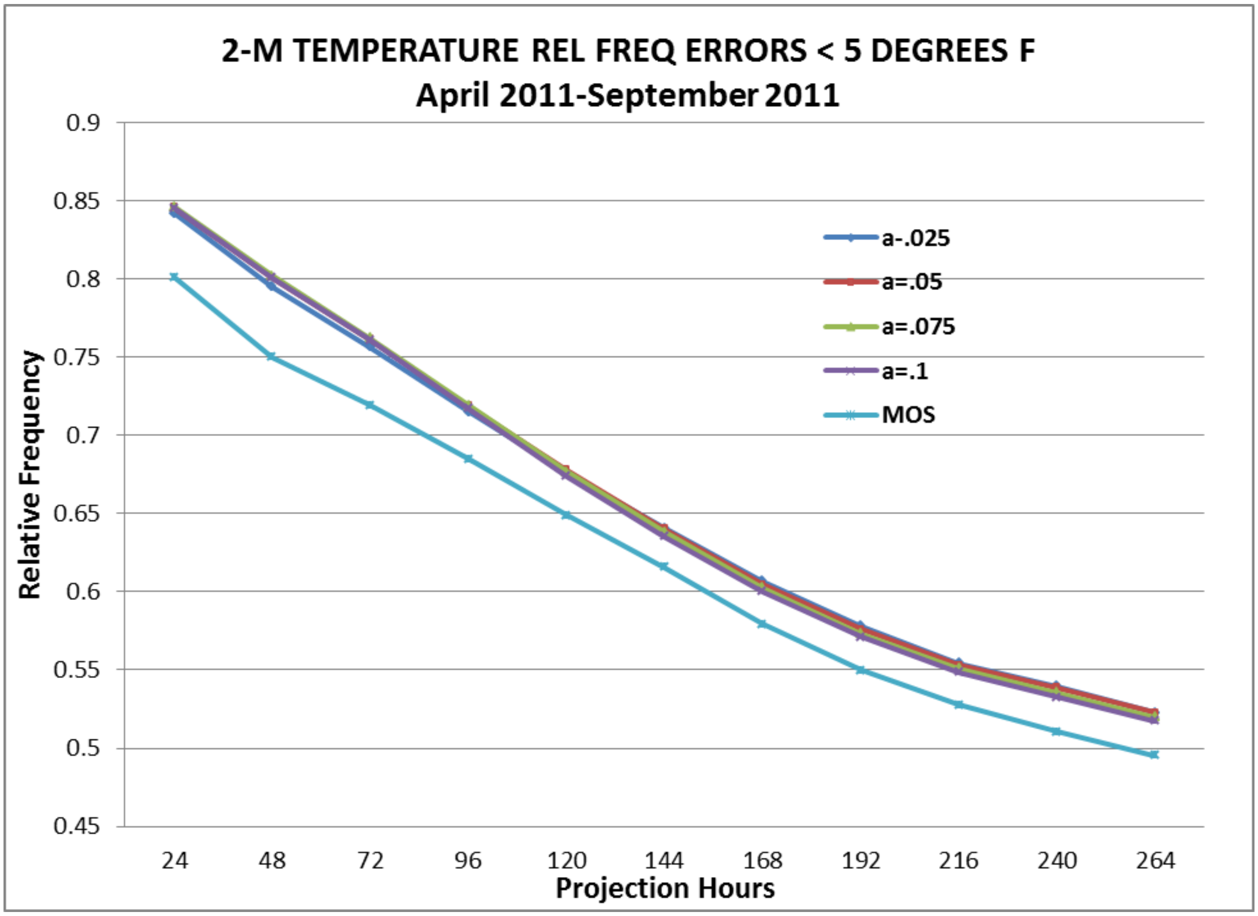
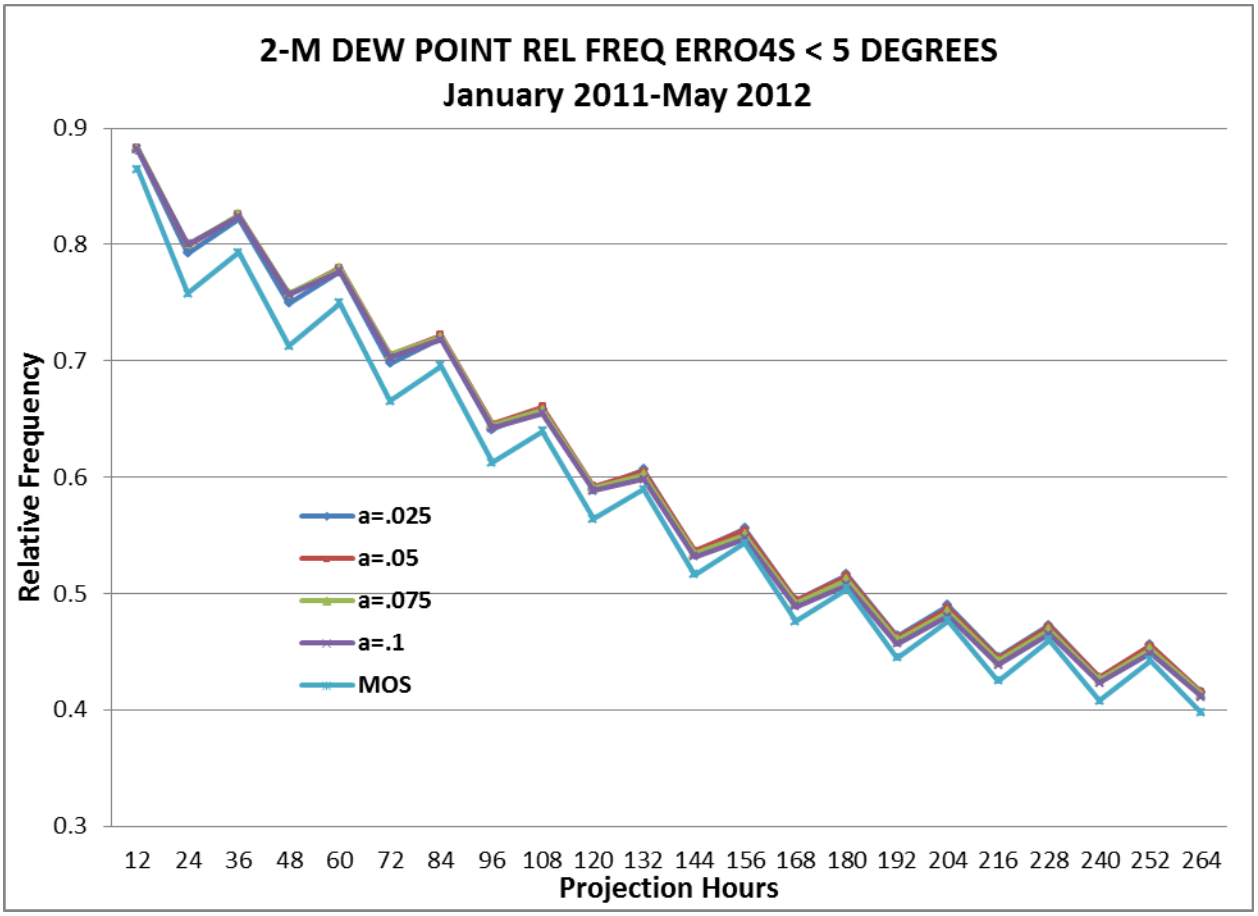


Figure 10. The same as Fig. 9, except for the 2012 warm season.



\

Figure 11. The same as Fig. 9, except for the warm and cool seasons combined for dew point.

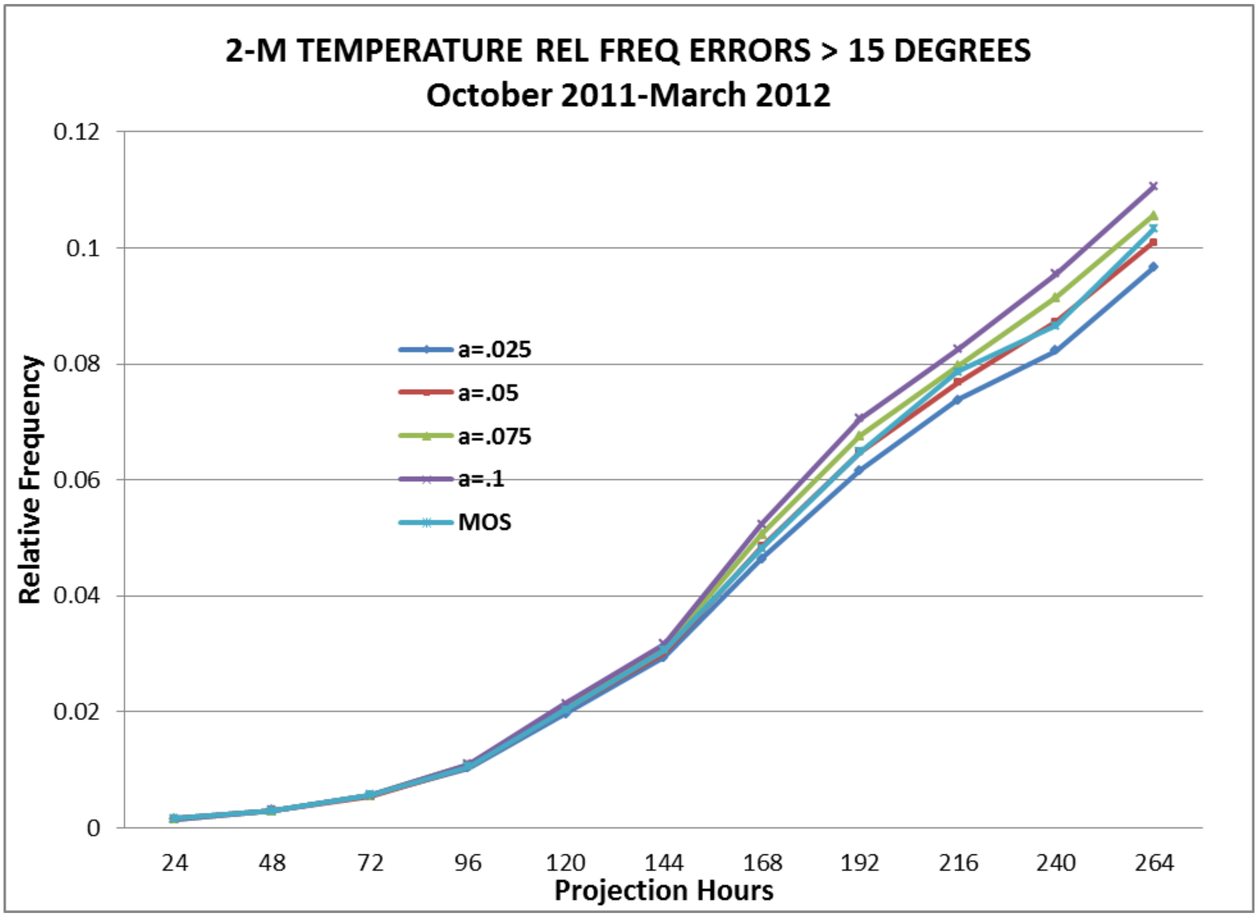


Figure 12. The percentage of temperature errors > 15 degrees F for the cool season.

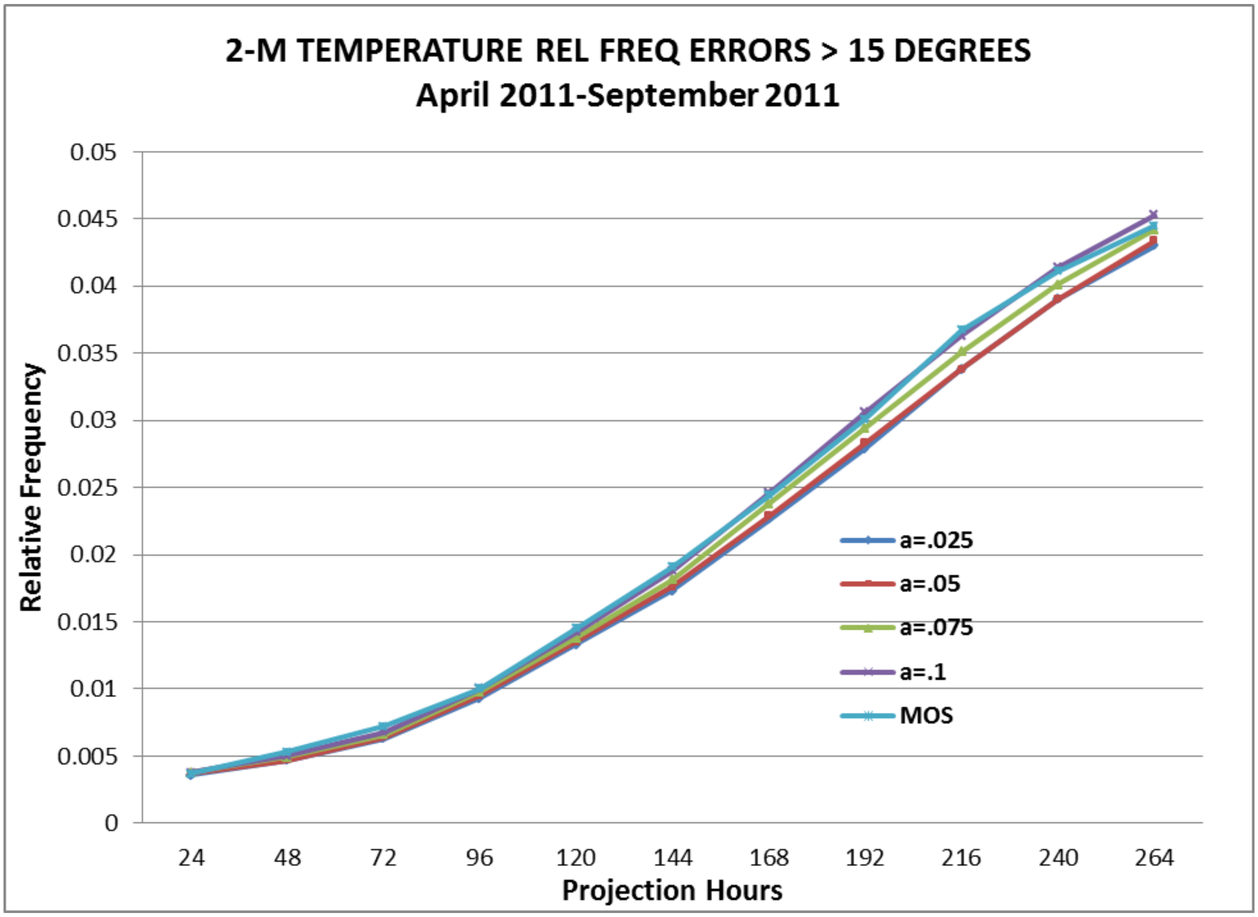


Figure 13. The same as Fig. 12, except for the 2012 warm season.

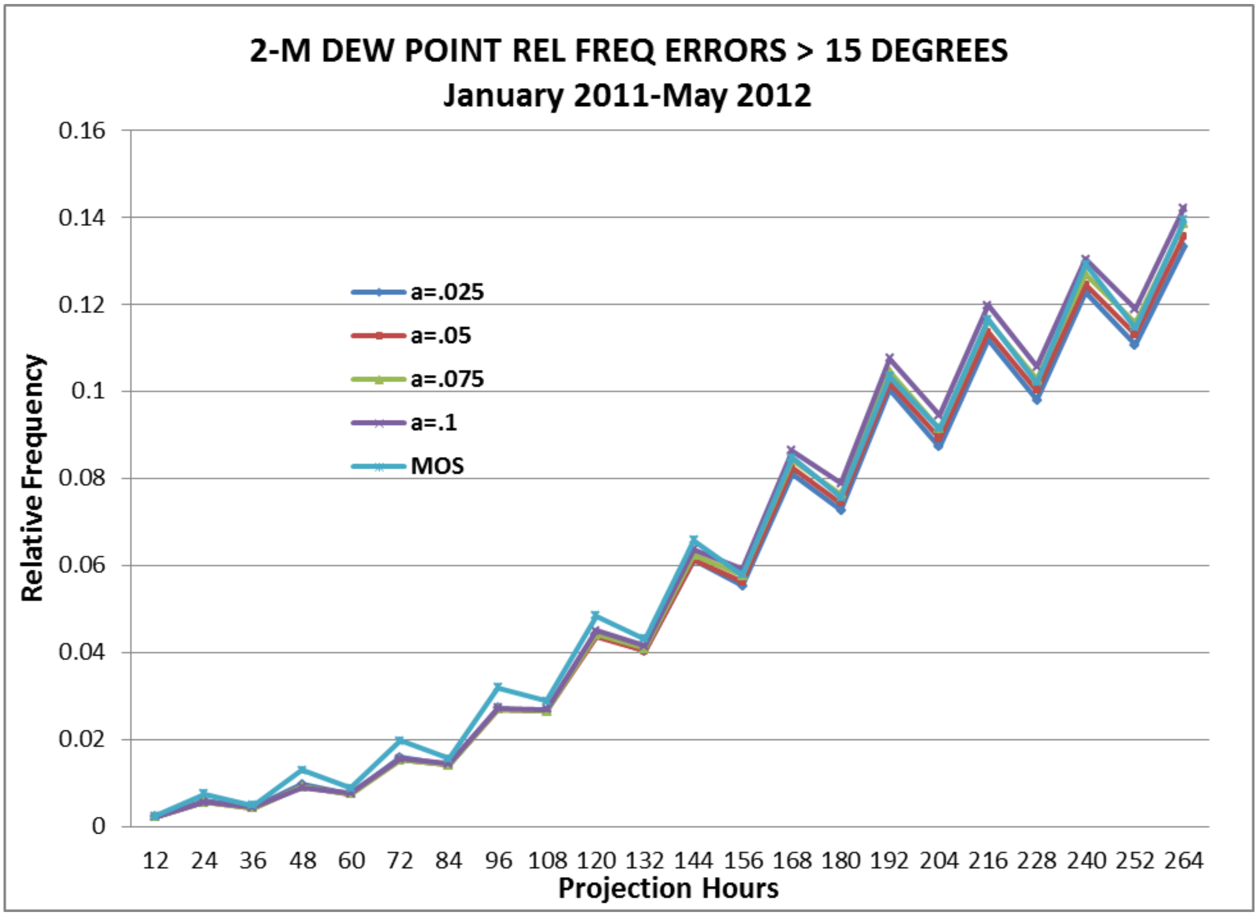


Figure 14. The same as Fig.12, except for the warm and cool seasons combined for dew point.

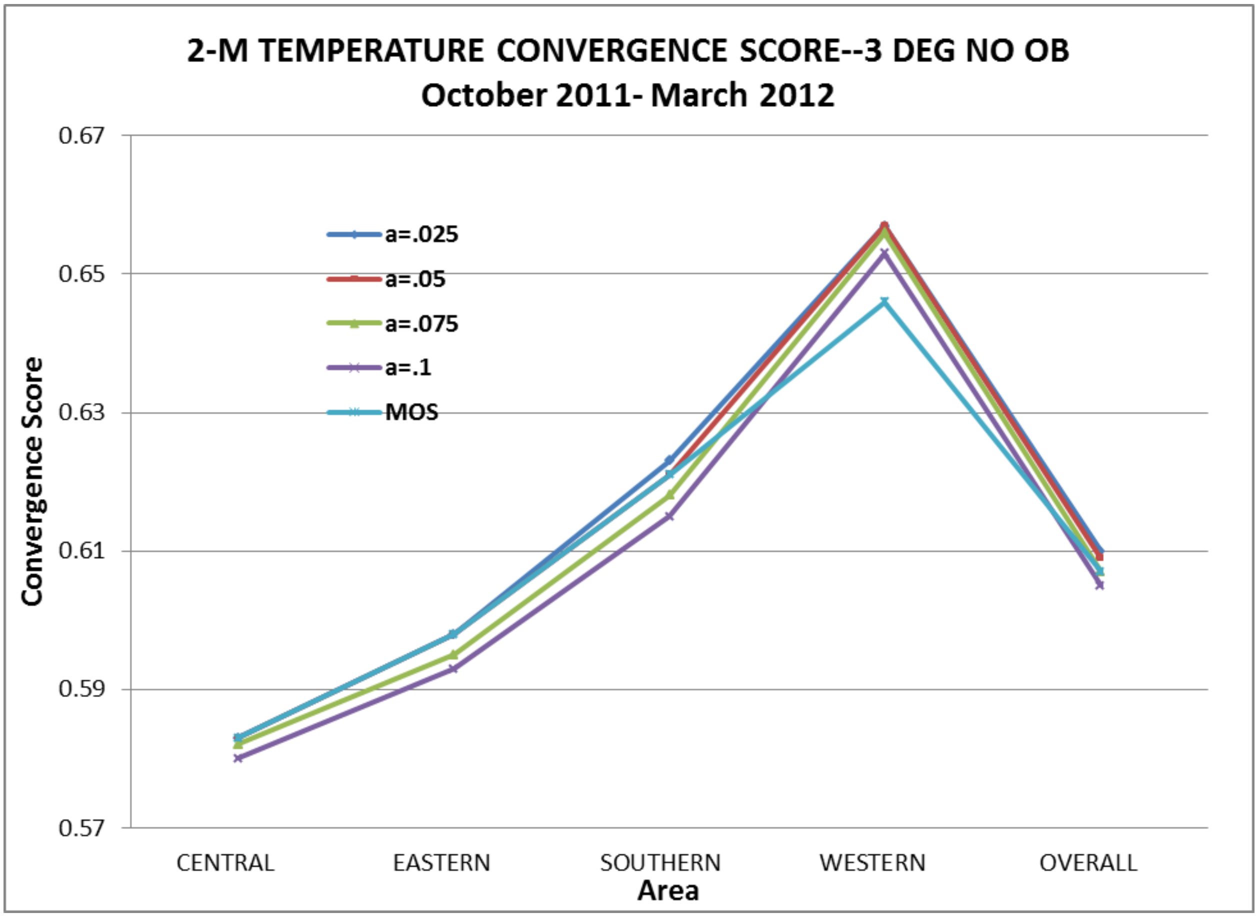


Figure 15. Convergence scores for the four NWS CONUS regions and overall for the cool season. The score for alpha = 0.025 is very close to that for alpha = 0.05, and can hardly be seen. The "3 deg no ob' refers to parameters for the score. Here, there is no penalty to the score if the change forecast is < 3 degrees F, and the observation to which the forecasts aspire is not used as a 0-h anchor point.

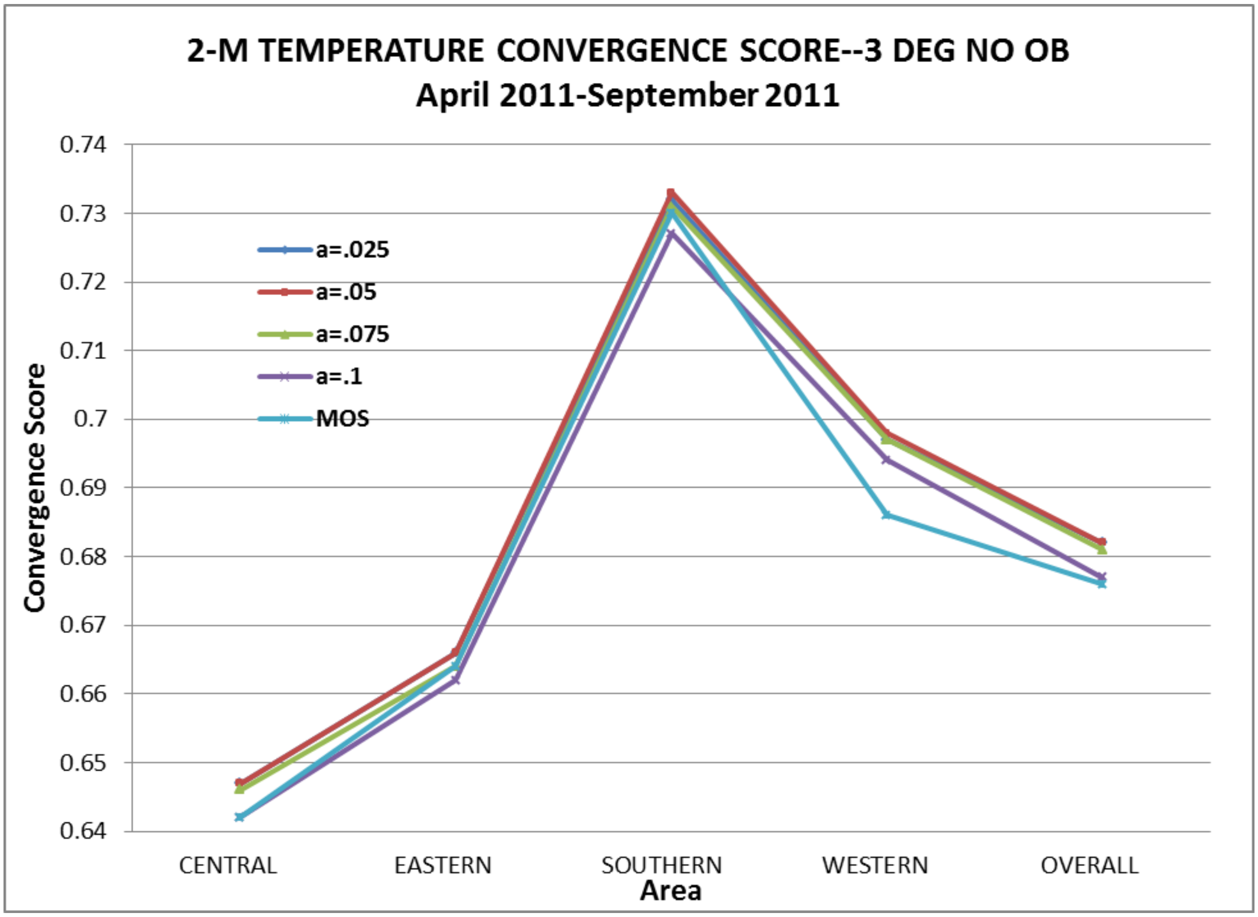


Figure 16. The same as Fig. 15, except for the warm season. The score for alpha = 0.025 is essentially the same as for alpha = 0.05, and can hardly be seen.

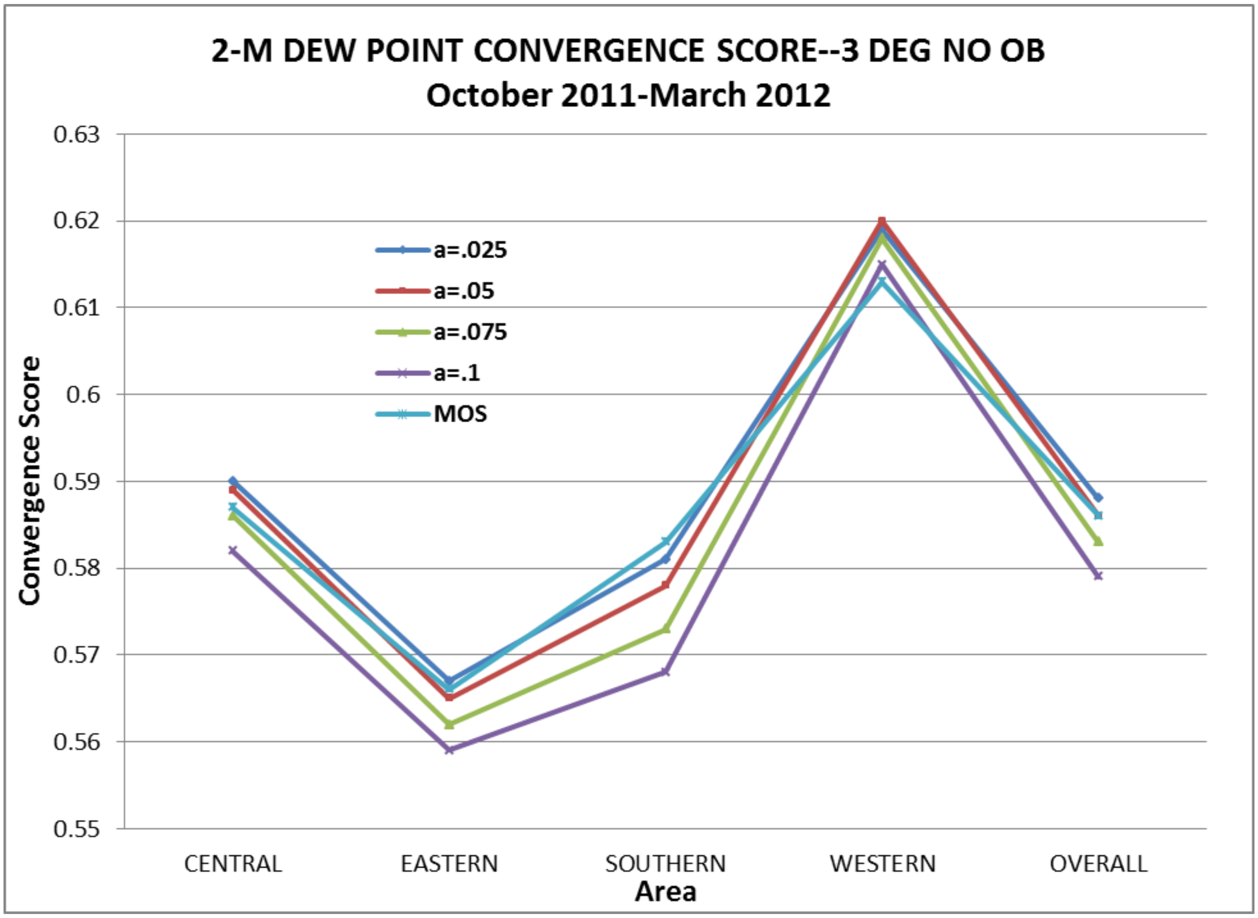


Figure 17. The same as Fig. 15, except for dew point for cool and warm seasons combined.

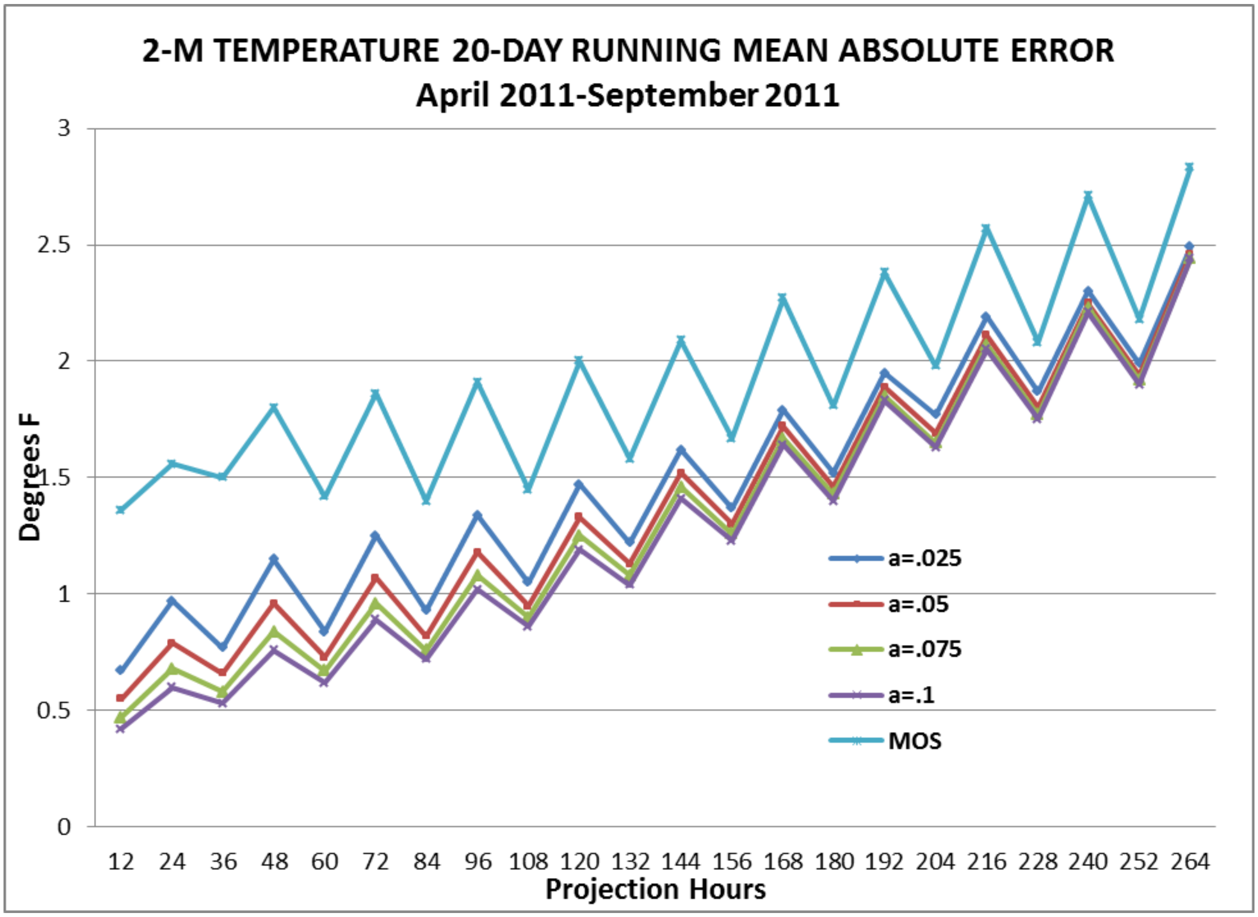


Figure 18. MAEs of forecasts over 20-day running periods for the warm season.

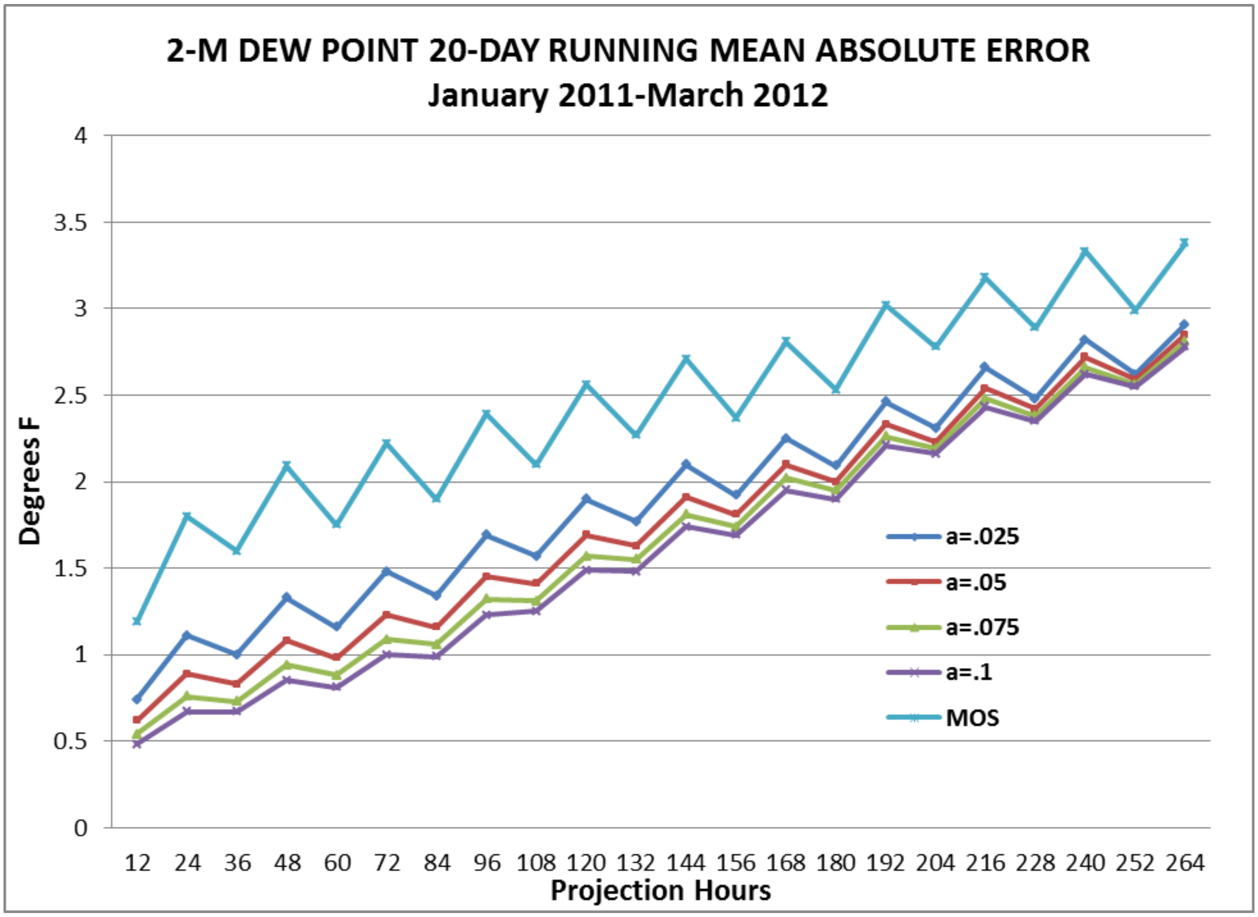


Figure 19. The same as Fig. 18, except for the cool and warm seasons combined for dew point.

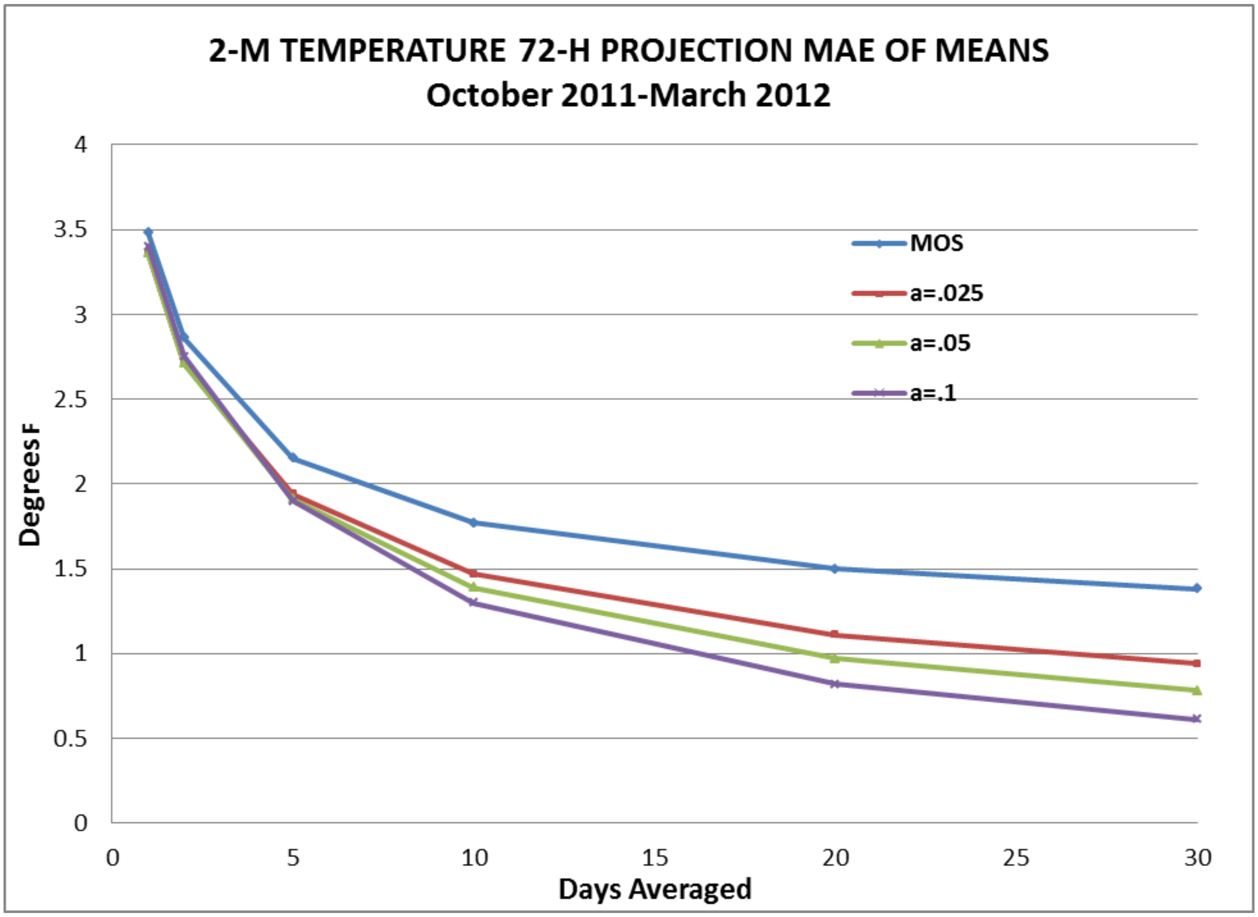


Figure 20. MAEs of 72-h forecasts as a function of averaging period (a = .075 is not shown).

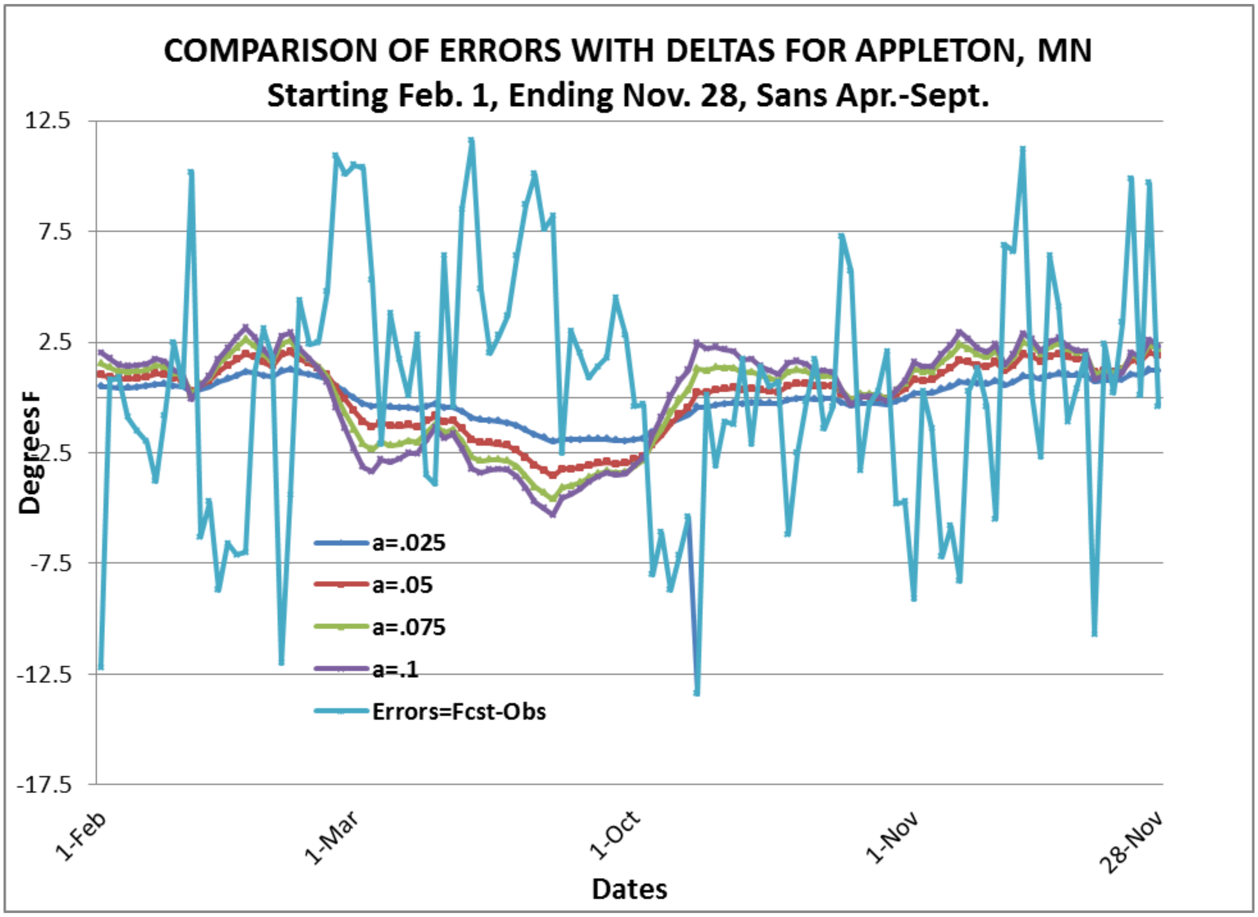


Figure 21. Errors in MOS temperature forecasts and the deltas associated with the four alphas tested.

APPENDIX

Implementation Considerations and Software

There are three aspects to the implementation of the bias correction method; they are not unique to the decaying average method. First, there is the process to implement the correction method, second is what to do about the seasonal equation changes on April 1 and October1, and lastly, how to deal with the objective analysis of the BC forecasts in U155. These three aspects are addressed below.

1. MAKING BIAS CORRECTED FORECASTS

Two existing MOS FORTRAN programs can be used for implementation—the same ones used in the testing reported here. The verification program U850 can be used in the usual way for comparison of MOS and BC MOS.

U201A

U201A is a slight modification to the well-known and heavily used U201. The modification was necessary to access forecasts made more than 99 hours in the past. The lookback feature RR in the MOS ID (see Glahn and Dallavalle 2000, Chapter 4) has only two digits, but three are needed to access forecasts made more than 99 hours ago. The subroutine GFETCH was modified a decade ago to accommodate a three digit number by annexing the digit previous to RR, the T, to the front of RR. This is signaled to GFETCH by setting the parameter ISTAV = 5555. Most programs do not take advantage of this capability, and a few changes were necessary to the U201 subroutines PRED21 and PRED22, as well as to U201 itself, the new version now being called U201A. U201A fulfills all functions of U201 and could be used with its name reverting back to U201 once sufficient regression tests have been made.

The inputs are:

-- Forecasts on one or more sequential files for a series of dates back in time to contain the forecasts verifying on the current date. Cycle runs (i.e., 0000, 0600, 1200, and 1800 UTC) are kept separate. These forecasts should be the ones before the temperature/dew point consistency checking is done.
-- Current observations on one or more sequential files.
-- External Random Access (ERA) file containing the last deltas for the all the projections of the forecasts to be bias corrected on Unit No. 49.

The outputs are:

-- Deltas on sequential files by date. This allows them to be used with forecasts to produce bias corrected forecasts for analysis and verification. The ID contains the projection of the forecasts with RR = 0, written with the current date/time NDATE.
-- The new deltas will overwrite the old ones on the ERA file on Unit No. 49. This ERA file will be used on the next cycle.

A new subroutine BDELT has hard coded IDs which currently pertain to consistency checked forecasts. These need to be changed to reflect unchecked forecasts, and it needs to be called from OPTION. It is unique, in that it writes directly to an ERA file. Also, note that the value of alpha is hard coded in BDELT, so once the specific value is decided on, it must be that value. The "cap" (see Section 5) is implemented in BDELT as Cap = 19.048 + .0794*tau.

Note that this U201A run can be made between cycles. That is, it can be scheduled any time before the next cycle 24 hours later. Because current observations are needed, elapsed time from the current run time is advantageous for all the observations to be obtained in operations.

For a cold start, to spin-up the process, U201A can be run for a series of dates, as it was for the testing reported here. The variable input, in addition to the usual CCCFFFBDD and tau will have a TRR that will go back to the cycle needed–the one that will match the hour of the dates being run. Note that tau and TRR will not be the same value, unless tau is evenly divisible by 24. In addition, each input variable must have a "9" in the G position of the ID. This allows certain manipulation, including merging T and RR, and also stripping off the TRR (and G) before writing the deltas. TRR and G are not required in the next phases of the process, so the use of these capabilities is self-contained and need not be worrisome. Also, when G = 9, the ERA file is closed when finished.

U720

As a postprocessing step, U720 can be run for the current cycle. This produces the bias corrected forecasts. U720 was not modified, but a new subroutine SBDEL is called from OPTION. The IDs are hard coded for consistency checked temperature and dew point; they will have to be changed for unchecked data.

The inputs are:

- Either a sequential file or an ERA file on Unit No. 48 of the uncorrected forecasts just produced, probably by U900 (the ERA file option has not been tested).
- An ERA file containing the deltas produced by U201 from the previous cycle 24 hours ago. If a cycle is missing, it won't matter, as there is no date is on the ERA file.`

The outputs are:

- A sequential and/or ERA file with the corrected forecasts.
- An ASCII file IP16 containing forecasts, deltas, and corrected forecasts. This can be used for analyzing the deltas and changes. If archived, it would show the corrections made in case of questions. With bias correction, it will be more difficult to reproduce how the "final" forecasts were arrived at. It can be done, but if this file is not saved, then the sequential file produced from the same run cycle of U201A would be needed. In short, something has to be archived if the production trail is to be reproduced.

Note that U720 just corrects the forecasts with the deltas; the deltas would not have to be arrived at through the decaying average method.

2. SEASONAL SPLICES

As indicated in the text, changeover from cool to warm season equations on April 1, and the other switch on October 1 already creates possible changes in MOS error characteristics on those dates. This could be made worse by bias correction without adequate planning for the deltas to use starting on those dates. The process outlined in Section 5 of the text whereby an overlap of 2 months is created in which both cool and warm season equations are evaluated and bias corrected, and a weighted average becomes the final forecast should eliminate the problem entirely. Because the making of the forecasts from the equations, and the postprocessing thereof, likely does not take a large amount of computer time, a human time-saving process would be to run both cool and warm season equations year around. Changeover could become completely automatic.

3. ANALYSIS OF BIAS CORRECTED FORECASTS

Currently, MOS forecasts are made without accessing mesonet data in real-time. If only the MOS forecasts are bias corrected at METAR stations, then a large percentage of the forecasts would not be bias corrected, and special accommodation would have to be made in the U155 analysis program. This could be done with the augmenting capabilities in use for other situations. With the spatial consistency in the larger MOS errors, augmentation should furnish a viable solution.

However, the mesonet data exist in NCEP buffer tanks. These could be accessed to bias correct most MOS forecasts. Note that the U201A run that needs them can be delayed from the current cycle by up to, say, 18 hours, giving plenty of time for the mesonet data to be received. This is a good option to pursue. Even so, a drift of a mesonet site's observation might cause "spots" in the analysis. Some added quality control process might have to be devised.

Dealing with the bias correction of mesonet stations for analysis purposes may well be the most problematic aspect of implementing bias correction.